

学校编码: 10384

分类号_____密级 _____

学号: 15420060153027

UDC _____

厦门大学

博 士 学 位 论 文

高频数据交易策略与波动性分析

Trading Strategy in High Frequency Financial Data and
Volatility Analysis

来 升 强

指导教师姓名: 朱 建 平 教授

专 业 名 称: 统 计 学

论文提交日期: 2 0 0 9 年 4 月

论文答辩日期: 2 0 0 9 年 月

学位授予日期: 2 0 0 9 年 月

答辩委员会主席: _____

评 阅 人: _____

2009 年 4 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为(高频数据交易策略与波动性分析)课题(组)的研究成果,获得(厦门大学优秀博士学位论文培育计划)课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

内容摘要

创立金融市场的原本目的是促进价格向价值的回归，但这种回归并不如预想的简单。由于市场内外各方力量对于价格 and 价值的认同并不一致，以至于金融市场波动性的产生。这就允许市场参与者无需掌握实体资本，仅仅通过分析价格走势，判断波动程度，就能获取丰厚利润。这大大刺激了金融市场的投机行为。而 21 世纪以来，现代金融市场中涌现了大量基于微观交易策略的投资方式。这种投资方式完全不同于 CAPM 所定义的投资组合，仅仅通过特定的技术指标和参数优化所得出的交易信号进行交易。这种技术分析色彩极强的交易策略长久以来都被学术界所质疑和批判，但技术分析仍然顽强的存活到现在，并在 1990 年代出现了复兴。这有两个方面的原因。

其一，在 1970 年代提出的 EMH 理论不断地受到质疑，大量的市场研究都证实金融市场的有效性只存在于数学模型之中。真实世界的投资者不仅不具备完全信息，其理性程度也是参差不齐的。代表性的重大发现应当是行为金融学中提出的反馈交易和羊群行为。前者指出投资者的非理性所导致的追涨杀跌行为，这一行为加重了股票价格的涨跌趋势、并延缓了价格恢复的过程。而后者指出，投资者群体性的非理性会导致股票价格与股票价值的长期偏离，在羊群行为突出的金融市场，市场对于资本的配置是极度低效率的。

其二，技术分析的争论促进了有关定量分析方法的产生。例如各种重抽样方法、滑动估计法、利用统计方法量化技术指标、交易指标选择中的后验知识判断等。虽然主流观点认为：在考虑了交易成本的情况下，技术分析基本无效。那为何技术分析为何没有消亡，反而广泛流传？进行了大量文献回顾后，我们发现了一些被主流研究忽略的但对于技术分析十分重要的地方：首先，以往的实证分析所选用的技术指标以过滤规则、移动平均法则为主。但仅凭这几类指标的实证结论就断言技术分析的无效，显得过于主观；其次，技术分析可以很灵敏的发现异常的价格波动，同时也会出现大量不显著的交易。对这些异常价格观测值在进行统计分析时，会破坏有关统计检验的分布假设。而大量不显著的交易会降低技术分析通过统计检验显著性；最后，技术分析强调的时效性被实证研究所忽略。本文所回顾的文献几乎都采用了逐日数据，甚至逐周数据。然而逐日数据仅提供了当天交易的信息汇总，本身并不能提供价格变化的过程信息，并不适合用于技术分析。

本研究选择了包括了 40 只全年正常交易的股票的逐笔交易数据，原始数据占用的硬盘空间超过 1.3G。可以说，本研究所采用的逐笔数据无论是从数据质量上、或是从数据规模上都是所查阅到的同类研究无法达到的。在完成了耗时一月之久的实证分析后，所得出的大量分析结论不仅具有很强的理论参考价值，对于微观交易策略制定也具有极强的实践指导意义。具体的，本研究的主要内容有如下一些。

(一) 通过遗传算法赋予交易策略更好的参数估计, 通过对交易策略组合的逐日收益率分析, 发现了一些微观交易策略的基本特征: (1) 交易策略对于价格波动具有很强的敏感性, 如果波动程度越强, 则获得高额收益的可能越大, 相反则无所获利; (2) 交易策略对于抽样频率很敏感。即使是同一种类型的交易策略, 应用在不同频率的交易价格数据上所得到的结果也相差很大; (3) 由于交易策略的高度并发性, 并不符合统计检验假设的前提条件, 但可以采用 VaR 方法从潜在损失的角度评价交易策略优劣。

(二) 以经典配对交易为代表的微观交易策略采用一定的规则选择配对资产, 不仅交易规则简单, 还可以同时对多项配对资产进行同时交易。强烈的实时价格驱动机制与平行交易的特点, 使得配对交易在获取巨额收益率的同时也会出现巨额的亏损, 可见对于微观交易策略, 保持流动性和风险控制显得尤为重要。而考虑到金融资产价格序列的特点, 本文提出对对数价格差建立协整关系, 并依照协整关系的显著性建立配对资产。这一交易策略的逐日收益率水平要优于经典方法。特别是在抽样频率为 30 分钟和 60 分钟时, 即使考虑交易成本, VaR 值与无交易成本的情形也十分接近。这表明协整关系在验证配对资产价差关系的稳定性和动态性上要优于经典的标准差方法。

(三) 结合套利定价理论, 本文重新建立了以收益率差余量的状态空间模型, 并采用卡尔曼滤波预测交易信号。所得到的组合逐日收益率不仅具有超高的异常值出现, 而且在各个抽样频率下都表现出了明显的正偏性。虽然会面临超过 20 倍的亏损, 但其最高 40 倍的获利能力不容小视, 大量的异常值的出现代表了交易策略对超额收益的追求, 而大量正异常值所导致的正偏性, 表明了基于滤波的配对交易策略的有效性。

(四) 对上述的各种微观交易策略进一步总结。在市场中性和套利定价理论的框架下, 重新剖析了交易策略的超额收益的来源。比较深入的讨论了基于市场中性的微观交易策略的含义和背后所蕴含的风险溢价。随后提出观点: 所谓市场中性其实不过是利用目前还没有被整个市场所发现或所考虑的个别风险因素进行套利。并建立了一个简明的动态模型解释交易策略对于真实的市场收益率波动性的影响机制, 很好的解释了由于选择投资方式的个体逐利行为导致了整个市场的真实波动性增加的现象。

关键词: 高频数据; 交易策略; 市场中性

Abstract

The purpose of running a financial market is to contribute to the regression of asset price to its value. Since various parts in and out of this market show their opinions with bids and asks, which lead to the chaos of prices. Among chaotic prices, professionals cultivate returns by predicting trends and estimating volatilities. And no capital asset is required in this business. Such a free lunch stimulates strong speculation in financial markets. From 21th century on, lots of trading strategy-based investment prevailed over financial market. This kind of investments is quite different from the definition of CAPM theory, it generates trading signals by propitiate technical indicators, so called technical analysis. This strategy has burden heavy critics from academia for decades. However technical analysis survived and even shew prosperity in the late 1990s. In this research, two reasons of many are given to explain such a prosperity.

On the one side, the famous EMH theory born in 1970s had been doubted for years, and hundreds of empirical analysis shew the fact that efficient market lives only in mathematical models. In the real world, investor is neither fully-informed nor homogeneous rationale. Two concrete discoveries from behavior finance are moment trading and herding behavior. The former tells us that investors who are lack of rationale, tend to follow the price and buy when price goes up, and sell when down. While the latter indicates that the irrational behavior at market level will force price divert from its value for quite a while, and hinder probable recursive process in a longer period. So in a market existing strong herding behaviors, the function of asset allocation is quite trivial.

On the other side, discussions against technical analysis lead to the generation of resampling methods, sliding estimation, statistical analysis-embedded technical analysis, bayesian judgement for choosing trading indicators, data selection techniques, and etc. However, the mainstream opinion still goes like “given the condition of transaction cost, technical analysis is inefficient” . Then following question appears hereby: if technical analysis is useless, then why it still lives with us and quite well? Is it because of defaults hidden in the opponents’ critics, or simply because of the stickiness of professionals and analysts? After a broad review, we find aspects which had been neglected by mainstream study but quite essential to the success of technical analysis: (1) most recent empirical research are based on filter rules, moving averages and etc, which are easy for programming and quite popular among analysts. But there are still other hundreds of technical indicators being used in the same time. So by testing the defaults of the few and alleging

all are not suitable and not reasonable as well. (2) technical indicator aims to capture variation in prices, not to explain it. In order to do well in predicting variations, more inferior predictions can be made. The extremes are never the favor of statistical tests, and a great number of insignificant trading signals can also lower the significant level. (3) the timing ability are essential to technical analysis, while it rarely appears on the formal research works. Among those empirical works reviewed, daily close price are mostly used, which obviously only supplies the result of daily trading, daily prices variations, which is the main source of interest for technical analysis, can be found in close price, plus some even use weekly data, whose conclusions maybe far from reality.

This study use a huge by-hand dataset over 1.3 gigabytes, containing each transaction records of 40 stocks in a year. Not a dataset used in all research works we reviewed, no matter in quality or in data scale, can be compared with ours. After one month's empirical study, results and conclusions show values for theoretical references, and it is quite valuable for practical designing of micro-level trading strategy as well. The main works done can be concluded as follows.

Firstly, we use genetic algorithm to give joint technical indicators a better estimation for their parameters. Some basic characteristics of trading strategy are found: (1) trading strategy is quite sensitive to the price variation, the greater the variation, the higher the gains; otherwise less or few. (2) trading strategy can be sensitive to the price sampling intervals. Given the same strategy, the different sampling frequency can generate many different results. (3) due to its parallel trading property, the behavior of its returns cannot be fitted in the traditional statistical analysis. Yet the value at risk method shows some potential to solve this problem.

Secondly, we use classical pairs trading to select pairing assets, which is quite easy and has good property of parallel trading. Strong price-driven mechanism and parallel trading, can give huge return and lost together. This reminds us that keeping an eye on liquidity and risk management is the key to success of trading strategy. Further considering the general attributes of financial times series, co-integration method is introduced in this study. After some modification, i.e., using the difference of logged prices instead of logged prices, we show that co-integration based pairs trading method outperforms classical one in many ways. Especially, in the sampling frequency of 30 and 60 minutes, even considering transaction costs, co-integration pairs trading does the same works as the performance of considering no transaction costs, and their VaRs are also close. So we concludes that co-integration methods shows better fitness for dynamical stability for

pairs assets.

Thirdly, under the arbitrage pricing theory(APT), we use the difference of returns to build a state space model. To make better prediction, kalman filter is incorporated. Results show that strategy-based portfolio's daily returns has extremely high observations, and positive skewness shows the potential of positive gains. Facing with 20 times lost, its 40 times gain is still lucrative. Again here we see, extremes represents eagerness of pursuing extreme profits.

Lastly, to give a more general view of these trading strategies discussed, we incorporate ideas of market neutral investing strategy and APT theory to decompose sources of returns of trading strategy. After an in-depth discussion about market neutral and risk premium, we conclude that: market neutral strategy is simply arbitrage over risks which is treated as unique risk or has not been considered by most investors. Additionally, we develop a easy and clear dynamic model to explain the mechanism that how strategy impacts market volatility. This model shows clearly that profit-pursuing investment-swinging behavior is one main source of volatility in financial market.

Keywords: High frequency data; Trading strategy; Market neutral

厦门大学博硕士

目 录

内容摘要	i
Abstract	iii
第一章 绪 论	1
第一节 交易策略的兴盛	1
第二节 现代金融理论的挑战	3
第三节 技术分析的演进	7
第四节 高频数据研究回顾	11
第五节 存在的问题	15
第六节 研究思路和创新	18
第二章 微观交易策略的影响因素	21
第一节 宏观经济因素	21
第二节 基本面因素	22
第三节 市场制度因素	23
第四节 市场因素	28
第三章 金融高频数据的统计特征	31
第一节 收益率描述分析	33
第二节 正态性检验	38
第三节 平稳性检验	39
第四节 序列相关检验	43
第五节 ARCH 效应检验	45
第六节 本章小结	45
第四章 基于遗传算法的微观交易策略	47
第一节 引言	47
第二节 遗传算法	48
第三节 技术指标与交易信号	54
第四节 实证分析	59
第五节 本章小结	69
第五章 经典的微观交易策略	71
第一节 配对交易起源	71
第二节 配对资产选择	72
第三节 收益率计算	73
第四节 实证分析	74
第五节 本章小结	81
第六章 基于协整关系的微观交易策略	83
第一节 相对定价	83
第二节 数据选择	84
第三节 协整检验	85
第四节 实证分析	86
第五节 本章小结	94

第七章 基于滤波方法的微观交易策略	95
第一节 套利定价与相对风险	95
第二节 随机收益率差模型	98
第三节 卡尔曼滤波	98
第四节 实证分析	104
第五节 本章小结	108
第八章 市场中性和市场波动性	109
第一节 引 言	109
第二节 市场中性与超额收益	110
第三节 市场中性与微观交易策略	112
第四节 市场真实波动性模型	114
第五节 实证分析	119
第六节 本章小结	124
第九章 总 结	125
附录A 最优卡尔曼增益阵的推导	127
附录B 部分 SAS 程序代码	129
参考文献	141
致 谢	147

CONTENTS

Chinese Abstract.....	i
Abstract.....	iii
Chapter 1 Introduction	1
1.1 The Prosperity of Trading Strategy	1
1.2 Challenges Against Modern Financial Theories	3
1.3 Evolution of Technical Analysis Research	7
1.4 Review of High-frequency Data Analysis.....	11
1.5 Existing Problems.....	15
1.6 Research Ideas and Originalities.....	18
Chapter 2 Impact Factors to Micro Trading Strategy	21
2.1 Macro Economic Factors	21
2.2 Fundamental Factors	22
2.3 Regulation and Mechanism Factors	23
2.4 Market Factors.....	28
Chapter 3 Statistical Characteristics of Financial High-frequency Data.....	31
3.1 Return Series Descriptive Statistics.....	33
3.2 Normal Tests	38
3.3 Stationary Tests	39
3.4 Serial Correlation Tests	43
3.5 ARCH Effect Tests	45
3.6 Summation.....	45
Chapter 4 GA-based Evolutionary Trading Strategy.....	47
4.1 Introduction	47
4.2 Genetic Algorithm	48
4.3 Technical Indicators and Trading Signals	54
4.4 Empirical Analysis	59
4.5 Summation.....	69
Chapter 5 Classical Pairs Trading Strategy	71
5.1 The Origin of Pairs Trading.....	71
5.2 Pairing Assets Selection	72
5.3 Computing Return Series	73
5.4 Empirical Analysis	74
5.5 Summation.....	81
Chapter 6 Cointegration-based Micro Trading Strategy	83
6.1 Relative Asset Pricing.....	83

6.2 Data Snooping	84
6.3 Cointegration Test	85
6.4 Empirical Analysis	86
6.5 Summation.....	94
Chapter 7 Filter-based Micro Trading Strategy.....	95
7.1 Arbitrage Pricing Theory and Relative Risk	95
7.2 Random Return Residual Model.....	98
7.3 Kalman Filter	98
7.4 Empirical Analysis	104
7.5 Summation.....	108
Chapter 8 Market Neutral and Market Volatility	109
8.1 Introduction	109
8.2 Market Neutral and Excessive Returns.....	110
8.3 Market Neutral and Micro Trading Strategy	112
8.4 Market Volatility Model	114
8.5 Empirical Analysis	119
8.6 Summation.....	124
Chapter 9 Conclusion	125
Appendix A Best Kalman Gain Derivation.....	127
Appendix B Main SAS Codes	129
Bibliography	141
Acknowledgement	147

第一章 绪 论

金融市场的主要目的是价格发现，并对经济资源进行有效配置。伴随着信息技术的高速发展，金融市场已经充分虚拟化，市场参与者仅通过网络连接就可以完成真实交易。在电子交易条件下，几乎所有的信息都是以电子形式传播。由于投资行为的不可重复性，能够在更短的时间里判断价格走势并形成交易信号，就意味着更大的获利空间和机会。

在学术界，越来越多的学者开始关注现代电子交易平台上市场均衡的形成机制和动态性，微观交易策略研究应运而生。对于实务界而言，微观交易策略是捕捉获利机会的手段；对于学术研究而言，它是竞争性市场均衡的实现工具。那么，微观交易策略是否有效，并具有实际可操作性？带着这一选题，本文展开了一系列基于微观交易策略的研究，并利用中国股票市场的逐笔数据进行了详尽的实证分析。

第一节 交易策略的兴盛

伴随着现代金融市场的发展，金融分析人员越来越多的依靠交易所提供的买卖交易记录来分析价格趋势，并建立投资决策。计算机辅助交易的优势就是在于能够实时响应市场变化，捕捉转瞬即逝的投资机会，并在极短的时间内下达交易指令，实现获利。一般地，计算机辅助交易系统由硬件系统、交易策略系统和交易分析系统三部分构成。主要包括了接收数据，选择最优交易策略模型、交易事前评价、指令集生成、交易后评价等步骤。交易系统的硬件配置要求并不高，一般的中小型服务器就能满足要求，网络传输带宽达到 50MBps 即可。目前，美国纽约股票交易所 (New York Stock Exchange, NYSE) 要求的网络传输带宽约 60MBps，而上海证券交易所富赢 (TopView) 行情系统要求网络带宽为 50MBps，数据刷新最大延时小于 3 秒^①。交易策略系统是由交易策略模型驱动，结合一定的执行策略，将交易策略模型发现的投资机会转化为一系列的买卖指令集，并由交易子系统直接发送至证券交易所的订单队列；而交易分析系统则负责算法交易策略的交易后分析和评价，并定期对历史数据进行回顾性检验。在这三个子系统中，交易策略系统中的交易策略设计尤为重要。由于交易策略设计涉及到大量计算机语言编程，因而计算机辅助交易也被称为算法交易 (algorithmic trading)。NYSE 称之为程序交易 (program trading)，并将其定义为“同时买入和卖出十五只股票以上、交易总额超过一百万美元的交易指令集”。计算机辅助交易系统还有许多别名，例如机器交易 (robot

^① 富赢 (TopView) 行情系统提供的数据主要包括三部分：(1) 分类账户统计，即将投资者账户分为机构、法人和个人三类，然后将沪市每只股票的持股账户进行分类统计；(2) 不同账户持股数的分档统计，共分 9 档 (1000 股以下、1000-1 万股、1 万-5 万股、5 万-10 万股、10 万-50 万股、50 万-100 万股、100 万-500 万股、500 万-1000 万股和 1000 万股以上) 进行分档；(3) 提供前十名最活跃席位的成交信息。考虑到 TopView 产品购买者能够随时查看机构投资者的交易情况，并大规模跟进。这种恶意的羊群行为能够迅速抬高目标股价，直接破坏了机构投资者的资产配置规划，众多基金公司表示强烈不满。目前，上海证券交易所已于 2009 年 1 月 1 日停用富赢行情系统

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕